Pairwise constraint propagation via low-rank matrix recovery

Zhenyong Fu¹ (✉)

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract As a kind of weaker supervisory information, pairwise constraints can be exploited to guide the data analysis process, such as data clustering. This paper formulates pairwise constraint propagation, which aims to predict the large quantity of unknown constraints from scarce known constraints, as a low-rank matrix recovery (LMR) problem. Although recent advances in transductive learning based on matrix completion can be directly adopted to solve this problem, our work intends to develop a more general low-rank matrix recovery solution for pairwise constraint propagation, which not only completes the unknown entries in the constraint matrix but also removes the noise from the data matrix. The problem can be effectively solved using an augmented Lagrange multiplier method. Experimental results on constrained clustering tasks based on the propagated pairwise constraints have shown that our method can obtain more stable results than state-of-the-art algorithms, and outperform them.

Keywords semi-supervised learning; pairwise constraint propagation; low-rank matrix recovery (LMR); constrained clustering; matrix completion

1 Introduction

Pairwise constraints provide prior knowledge as to whether two data points belong to the same class or not, known as must-link constraints and cannot-link constraints, respectively. Generally, we cannot infer instance labels from only pairwise constraints, especially for multi-class data. This

means that pairwise constraints are weaker and thus more general than explicit labels on data. Pairwise constraints have been widely used in the context of clustering with side information, also called semi-supervised clustering or constrained clustering, where it has been shown that the presence of appropriate pairwise constraints can often improve the performance [1–5]. While it is possible to infer pairwise constraints from domain knowledge or user feedback, in practice, the availability of such constraints is scarce. Pairwise constraint propagation aims to produce a large quantity of pairwise constraints from scarce known constraints.

One way to utilize pairwise constraints for constrained clustering is to trivially set the similarities between the constrained data to 1 and 0 for must-link and cannot-link constraints, respectively [3]. This approach only adjusts the similarities between constrained data, without propagating the constraint information to other data. In contrast, much research shows that it is more effective to propagate known constraint information to other unconstrained data [6–8]. In Ref. [9], the pairwise constraints are propagated to unconstrained data using a Gaussian process. However, as noted in Ref. [9], this method makes certain assumptions about constraint propagation, especially with respect to two-class problems.

Recently, some research has tried to solve the problem of pairwise constraint propagation under the semi-supervised learning framework that is usually based on the manifold assumption (objects on the same manifold structure tend to have similar pairwise constraint settings) [10–15]. For instance, in Ref. [7], Li et al. applied the manifold assumption inversely. They proposed to learn a mapping with known pairwise constraints so that two must-link objects are mapped to the same manifold while

1 College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China. E-mail: fuzy@njupt.edu.cn (✉).

Manuscript received: 2014-12-05; accepted: 2015-03-18

two cannot-link objects are mapped to different manifolds. Propagation of the constraint relationship is achieved by making the mapping smooth on the manifolds. Li et al.'s method can be formulated as a semidefinite programming (SDP) problem. Although this method can tackle multi-class problems, current SDP solvers are severely limited in the size of problems they can solve. Other work also based on the manifold assumption is the exhaustive and efficient pairwise constraint propagation (E²CP) method proposed by Lu and Ip [8]. It decomposes the constraint propagation problem into a series of two-class label propagation subproblems, each of which can be solved through semi-supervised learning based on k -nearest neighbor graphs. The basic underpinning of Lu and Ip's method is still the manifold assumption, i.e., that similar objects should have similar propagation results.

However, the manifold assumption only concerns local structure. In other words, it does not concern those objects that are not close each other or do not share the same manifold structure. In this paper, we make attempt to capture global structure in the constraint propagation problem. We observe that the constraint matrix, whose entries are the propagated pairwise constraints, should be globally low-rank, which has not been noted in previous work on constraint propagation. Therefore, we address constraint propagation via the low-rank assumption, i.e., among all of the solutions for constraint propagation, we aim to choose the lowest-rank one.

With the low-rank assumption, we formulate pairwise constraint propagation as a low-rank matrix recovery (LMR) problem. In our work, we first present a matrix completion formulation, which is a special and simpler case of the low-rank matrix recovery problem, for constraint propagation. Our experiments show that this matrix completion method works well in practice as long as there are enough known pairwise constraints. However, as the available pairwise constraints decrease, the performance of such a simple matrix completion method degrades very quickly. This indicates that when lacking sufficient constraints, to improve the propagation performance, we need some other information. Recent progress in transductive learning addresses the semi-supervised learning problem using a matrix completion framework [16,

17]. The main novelty of these methods is to make the data matrix and the label matrix jointly low-rank. In the following, we will refer to such methods as joint-MC. Although joint-MC methods can be directly applied to solve the constraint propagation problem, these methods neglect the influence of noise in the data. In this paper, we adopt a joint low-rank approach for constraint propagation, but use a more general low-rank matrix recovery method which explicitly deals with the data noise. More concretely, joint-MC methods assume that the noisy data matrix and the constraint matrix should be jointly low-rank, while our method assumes that the clean data matrix and the constraint matrix should be jointly low-rank. Our experimental results show that the latter assumption is more reasonable. The proposed method is formulated as a low-rank matrix recovery problem, which can be effectively solved with an augmented Lagrange multiplier algorithm.

For comparison with the existing propagation approaches, we use the propagated constraints obtained by our method to adjust the similarity between data points, as do previous works, so that they can be incorporated into the data clustering process. To evaluate the effectiveness of the proposed method, we select two real-life image data sets for constrained clustering tasks. Our experimental results show that the proposed method can achieve more stable and better performance than the state-of-the-art.

The main contributions of this paper can be summarized as follows:

- We address the constraint propagation problem on the basis of the low-rank nature of this problem, which has not been noted by previous work.
- Our work shows that when there are enough known pairwise constraints, the simplest matrix completion formulation is very effective for constraint propagation.
- We further propose a more general low-rank matrix recovery solution for constraint propagation that works well even with insufficient pairwise constraints.

The remainder of this paper is organized as follows. Section 2 presents our low-rank matrix recovery algorithm for constraint propagation. Section 3 explains how to apply the proposed method to

constrained clustering tasks. In Section 4, our methods are evaluated on two real-life image data sets. Finally, Section 5 gives the conclusions drawn from the experimental results.

2 Pairwise constraint propagation via LMR

2.1 Low-rank structure in constraint propagation

Given a data set of n objects $\mathcal{X} = \{x_1, \dots, x_n\}$ and two sets of pairwise constraints, denoted respectively by $\mathcal{M} = \{(x_i, x_j)\}$ where x_i and x_j should be in the same class and $\mathcal{C} = \{(x_i, x_j)\}$ where x_i and x_j should be in different classes, the goal of pairwise constraint propagation is to propagate these pairwise constraints across the entire data set.

To propagate the initial pairwise constraints on $\mathcal{X}$, we first represent the two types of pairwise constraints, i.e., $\mathcal{M}$ and $\mathcal{C}$, using a single constraint matrix $Y = \{Y_{ij}\}_{n \times n}$:

$$Y_{ij} = \begin{cases} +1, & (x_i, x_j) \in \mathcal{M} \\ -1, & (x_i, x_j) \in \mathcal{C} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Given $\mathcal{X}$ and Y , the problem of pairwise constraint propagation is to construct a matrix $F = \{F_{ij}\}_{n \times n}$ so that for any pair instances $(x_i, x_j) \in \mathcal{X} \times \mathcal{X}$, the element F_{ij} can be used to predict the pairwise constraint relationship between x_i and x_j , i.e., $F_{ij} > 0$ means that x_i and x_j should be must-link while $F_{ij} < 0$ means that x_i and x_j should be cannot-link.

As mentioned above, one effective method to solve this problem is to view the constraint propagation as a two-class classification problem, where the “positive class” is the must-link relationship and the “negative class” is the cannot-link relationship [8, 14]. However, in this paper, we view constraint propagation as a low-rank matrix recovery problem.

As shown in Fig. 1, in the problem of pairwise constraint propagation, we have a partially observed constraint matrix Y and wish to obtain the whole constraint matrix, that is, we have to complete the missing entries in the constraint matrix. This process satisfies the requirement of matrix completion. More importantly, we notice that the constraint matrix has low-rank structure. In fact, each row (and column) of

Fig. 1 Red and blue respectively represent +1 (a must-link constraint) and -1 (a cannot-link constraint) in the constraint matrix. Constraint propagation can be viewed as a process that recovers the underlying low-rank matrix from partial observation.

the constraint matrix can be viewed as the constraint vector of an object in the data set. It can be observed that in a desired constraint matrix, objects from the same class should have the same constraint vector. Thus, the matrix rows (and columns) from the objects that belong to a given special class form a rank-1 subspace and the total rank of the desired constraint matrix should equal the number of classes in the data set. In summary, we can view constraint propagation as a low-rank matrix completion problem, which can be formulated as the following optimization problem:

$$\begin{aligned} \min_F \quad & \text{rank}(F) \\ \text{s.t.} \quad & F_{ij} = Y_{ij}, \quad \forall (x_i, x_j) \in \mathcal{M} \cup \mathcal{C} \end{aligned} \quad (2)$$

The above matrix completion problem is difficult to solve due to the discrete nature of the rank function. As suggested by Refs. [18, 19], Problem (2) can be replaced by solving the following convex surrogate:

$$\begin{aligned} \min_F \quad & \|F\|_* \\ \text{s.t.} \quad & F_{ij} = Y_{ij}, \quad \forall (x_i, x_j) \in \mathcal{M} \cup \mathcal{C} \end{aligned} \quad (3)$$

Here, $\|\cdot\|_*$ denotes the nuclear norm [20] of a matrix, i.e., the sum of the singular values of the matrix.

Problem (3) is the simplest matrix completion formulation for constraint propagation. In matrix completion research [19, 21], Problem (3) has been recently carefully studied, and can be solved by singular value thresholding (SVT) [22], fixed point continuation (FPC) [23], or augmented Lagrange multiplier (ALM) [24]. Figure 2 shows a clustering performance comparison between simple-MC, E²CP, and NCuts on the Scene data set (the experimental details will be described in Sections 3 and 4). It can be seen that when the known

Fig. 2 Clustering performance comparison between NCuts, E²CP, and simple-MC on the Scene data set.

pairwise constraints are fewer than 0.5%, simple-MC even does not perform better than NCuts, which is effectively a spectral clustering algorithm but without considering pairwise constraints. On the other hand, when the known constraints are fewer than 1%, the performance of simple-MC is still worse than that of E²CP, which is the baseline constraint propagation approach. Such results verify the conclusion that we can exactly recover a low-rank matrix only when we have enough observed matrix entries [19], i.e., known initial pairwise constraints in the context of constraint propagation.

2.2 Low-rank matrix recovery formulation for constraint propagation

Since the known initial pairwise constraints are usually sparse, to further improve the performance of the low-rank matrix recovery based propagation method, we have to resort to some auxiliary information. Joint-MC methods from the area of transductive learning assume that the data matrix and the label matrix should be jointly low-rank [16, 17]. Although these matrix completion based transduction learning algorithms can be directly used to solve constraint propagation, as discussed earlier, joint-MC methods may be unreliable in the presence of unclean data due to the lack of robustness to noise. In this section, we attempt to develop a general low-rank matrix recovery solution for constraint propagation.

Let D denote the data matrix. Joint-MC methods consider that the matrix (D, F) should be jointly low-rank. In this paper, we assume that the data matrix D is the sum of a clean data matrix A and an additive error matrix E , i.e., $D = A + E$. Unlike

the joint-MC methods, we intend to make the matrix (A, F) , which concatenates the clean data matrix A and the propagated constraint matrix F , jointly low-rank. This idea can be formulated as the following optimization problem:

$$\begin{aligned} \min_{F, A, E} \quad & \text{rank}((A, F)) \\ \text{s.t.} \quad & D = A + E \\ & F_{ij} = Y_{ij}, \quad \forall (x_i, x_j) \in \mathcal{M} \cup \mathcal{C} \end{aligned} \quad (4)$$

There are two issues with Formulation (4). As in Problem (2), $\text{rank}()$ is a non-convex function and difficult to optimize. We can relax the rank function to the convex nuclear norm $\|\cdot\|_*$. In addition, the discrepancy between A and D , i.e., the noise matrix E , should be minimized. These two issues lead to the following convex optimization problem for constraint propagation:

$$\begin{aligned} \min_{F, A, E} \quad & \|(A, F)\|_* + \lambda \|E\|_1 \\ \text{s.t.} \quad & D = A + E \\ & F_{ij} = Y_{ij}, \quad \forall (x_i, x_j) \in \mathcal{M} \cup \mathcal{C} \end{aligned} \quad (5)$$

where λ is a positive weighting parameter. In Formulation (5), we use the ℓ_1 norm (the sum of the absolute values of the matrix entries) to measure the additive error matrix E , because the ℓ_1 norm can handle arbitrarily large entries in E . The ℓ_1 norm works well in our experiments. In addition, the extension of the proposed method from the ℓ_1 norm to other matrix norms, such as the Frobenius norm or $\ell_{2,1}$ norm [25] is straightforward.

It should be noted that there are three considerations in Formulation (5). On one hand, we wish to remove the noise from a submatrix D in the joint matrix (D, F) . On the other hand, we wish to complete the other submatrix F in the joint matrix (D, F) from the partially observed matrix Y . Finally, we wish the new joint matrix (A, F) to be low-rank. This differs from the formulations of robust principal component analysis (RPCA) [26] or matrix completion [19]. In fact, Formulation (5) can be viewed as a combination of RPCA and matrix completion, specially designed for the problem of pairwise constraint propagation.

2.3 Solving the optimization problem

In this section, we apply the augmented Lagrange multiplier (ALM) method to solve Formulation

(5). We first convert Problem (5) to the following equivalent problem:

$$\begin{aligned} \min_{F,A,E} \quad & \|(A, F)\|_* + \lambda \|E\|_1 \\ \text{s.t.} \quad & A + E = D \\ & F + R = Y, \quad \pi_\Omega(R) = 0 \end{aligned} \quad (6)$$

where $\Omega = \mathcal{M} \cup \mathcal{C}$ and $\pi_\Omega : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ is a linear operator that keeps the entries in Ω unchanged and sets those outside Ω (i.e., in $\bar{\Omega}$) to zero. Following recent work in low-rank matrix recovery [24], in Problem (6), we use the matrix R to compensate for the unknown entries of Y . Let $B = (A, F)$, $M = (E, R)$, and $N = (D, Y)$. Problem (6) can be reformulated as follows:

$$\begin{aligned} \min_{F,A,E} \quad & \|B\|_* + \lambda \|E\|_1 \\ \text{s.t.} \quad & B + M = N, \quad \pi_\Omega(R) = 0 \end{aligned} \quad (7)$$

Problem (7) can be solved by minimizing the following augmented Lagrangian function:

$$\begin{aligned} L(B, E, R, H, \mu) = & \|B\|_* + \lambda \|E\|_1 \\ & + \text{tr}[H^T(N - B - M)] + \frac{\mu}{2} \|N - B - M\|_F^2 \end{aligned} \quad (8)$$

where $\mu > 0$ is a penalty parameter and $H = (H_e, H_r)$ is the Lagrangian multiplier, in which the submatrices H_e and H_r are the Lagrange multipliers that correspond to the constraints $A + E = D$ and $F + R = Y$ respectively. The above problem can be solved either by exact or inexact ALM algorithms, as suggested in Ref. [24]. For efficiency, we choose the inexact ALM, in which in each iteration we need to alternatively update the matrices B , E , R , and H .

The following soft-thresholding (shrinkage) operator is used in the process of solving Problem (7):

$$\mathcal{S}_\epsilon[x] \doteq \begin{cases} x - \epsilon, & \text{if } x > \epsilon \\ x + \epsilon, & \text{if } x < -\epsilon \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

The update to matrix B optimizes the following subproblem:

$$\begin{aligned} \min_B \quad & L(B, E_k, R_k, H_k, \mu_k) \\ \Leftrightarrow \min_B \quad & \frac{1}{\mu_k} \|B\|_* + \frac{1}{2} \left\| B - \left(N - M_k + \frac{1}{\mu_k} H_k \right) \right\|_F^2 \end{aligned}$$

As described in Ref. [22], the above subproblem for matrix B has a closed-form solution: $B_{k+1} = U \mathcal{S}_{\mu_k^{-1}}[S] V^T$, where $(U, S, V) = \text{svd}(N - M_k +$

$\mu_k^{-1} H_k)$. The update to matrix E solves the following subproblem:

$$\begin{aligned} \min_E \quad & L(B_{k+1}, E, R_k, H_k, \mu_k) \\ \Leftrightarrow \min_E \quad & \frac{\lambda}{\mu_k} \|E\|_1 + \frac{1}{2} \left\| E - \left(D - A_{k+1} + \frac{1}{\mu_k} H_{e,k} \right) \right\|_F^2 \end{aligned}$$

According to the well-known analysis in Ref. [27], the subproblem for matrix E also has a closed-form solution: $E_{k+1} = \mathcal{S}_{\lambda \mu_k^{-1}}(D - A_{k+1} + \mu_k^{-1} H_{e,k})$. The update to matrix R solves the following subproblem:

$$\begin{aligned} \min_R \quad & L(B_{k+1}, E_{k+1}, R, H_k, \mu_k) \\ \Leftrightarrow \min_R \quad & \frac{1}{2} \|R\|_F^2 - \text{tr} \left[R^T \left(Y - F_{k+1} + \frac{1}{\mu_k} H_{r,k} \right) \right] \end{aligned}$$

The above subproblem for matrix R is a quadratic optimization problem with closed-form solution: $R_{k+1} = \pi_{\bar{\Omega}}(Y - F_{k+1} + \mu_k^{-1} H_{r,k})$, where the constraint $\pi_\Omega(R) = 0$ is enforced. In addition, following the general framework of ALM, the Lagrangian multiplier H is updated as follows:

$$H_{k+1} = H_k + \mu_k (N - B_{k+1} - M_{k+1})$$

and the parameter μ is updated via $\mu_{k+1} = \rho \mu_k$ with $\rho > 1$ (see Ref. [28] for further details). The convergence properties of the inexact ALM have been proved in Ref. [24].

Once we have the converged matrix $B_k = (A_k, F_k)$, we can extract the submatrix F_k from B_k and set $F = F_k$ as the constraint propagation result. We summarise the proposed method in Algorithm 1, which we call the low-rank matrix recovery based pairwise constraint propagation (LMRPCP).

Algorithm 1: Low-rank matrix recovery based pairwise constraint propagation

Input: Data matrix D , observed constraint matrix Y .

Initialize: $B = 0$, $E = 0$, $R = 0$, $H = 0$, $\mu = 1\text{e} - 7$,
 $\lambda = 1\text{e} - 3$, $\rho = 1.2$, $\epsilon = 1\text{e} - 8$.

while not converged **do**

- 1) $(U, S, V) = \text{svd}(N - M_k + \mu_k^{-1} H_k)$,
- 2) $B_{k+1} = U \mathcal{S}_{\mu_k^{-1}}[S] V^T$.
- 3) $E_{k+1} = \mathcal{S}_{\lambda \mu_k^{-1}}(D - A_{k+1} + \mu_k^{-1} H_{e,k})$
- 4) $R_{k+1} = \pi_{\bar{\Omega}}(Y - F_{k+1} + \mu_k^{-1} H_{r,k})$
- 5) $H_{k+1} = H_k + \mu_k (N - B_{k+1} - M_{k+1})$
- 6) $\mu_{k+1} = \rho \mu_k$
- 7) $k = k + 1$
- 8) Check the convergence condition:
 $\|N - B_k - M_k\|_\infty < \epsilon$

end while

Output: $F = F_k$

3 Application to constrained clustering

In previous work, the propagation result F is further used to improve the clustering algorithm. For fair comparison, we use spectral clustering as the basic clustering algorithm. As in Ref. [8], we incorporate F into the spectral clustering algorithm by adjusting the similarities between the data points according to F using the following adjustment method:

$$w_{ij}^* = \begin{cases} 1 - (1 - F_{ij})(1 - w_{ij}), & F_{ij} \geq 0 \\ (1 + F_{ij})w_{ij}, & F_{ij} < 0 \end{cases} \quad (10)$$

where $W = [w_{ij}]_{n \times n}$ is the similarity matrix defined on the data set. It can be seen that the above adjustment will increase the similarity when $F_{ij} > 0$ and decrease it when $F_{ij} < 0$. Let $W^* = [w_{ij}^*]_{n \times n}$ be the adjusted similarity matrix. We apply the spectral clustering algorithm [29] to W^* to form the final clusters.

4 Experimental results

In this section, we apply the propagated pairwise constraints to constrained clustering tasks and test the performance of the proposed method on two real-life image data sets. For comparison, the results of three well-known, most closely related algorithms are also reported: Lu and Carreira-Perpinán's affinity propagation [9], Lu and Ip's exhaustive and efficient constraint propagation [8], and Goldberg et al.'s joint-MC [17]. We use normalized cuts (NCuts) [30], which is effectively a spectral clustering algorithm but without considering pairwise constraints, as the baseline.

We first describe the experimental setup, including similarity matrix computation, the performance measure, and parameter selection. Then we compare the proposed algorithm to the other four approaches on two image data sets.

4.1 Experimental setup

We tested the proposed algorithm on two different image data sets. The first one contained eight scene categories from MIT [31], including four man-made scenes and four natural scenes. The total number of images was 2688. The size of each image in this Scene data set was 256×256 pixels. The second data set contained images from a Corel collection. We selected 15 categories including bus, sunrise/sunset, plane, foxes, horses, coins, gardens, eagles, models,

sailing, stream trains, racing car, pumpkins, rockies, and fields. Each of these categories contained 100 images. Therefore, this selected Corel data set had a total of 1500 images. The size of each image in this data set was 384×256 or 256×384 pixels. Some image examples from these two data sets are given in Fig. 3.

For these two image data sets, we chose two different feature sets as used in Refs. [32] and [33], respectively. Following Ref. [32], SIFT descriptors were used for the Scene data set, while, following Ref. [33], joint color and Gabor features were used for the Corel data set. These features were chosen to ensure a fair comparison to state-of-the-art techniques. More concretely, for the Scene data set, we extracted SIFT descriptors for 16×16 pixel blocks computed over a regular grid with spacing of 8 pixels. For the Corel data set, we divided each image into blocks of 16×16 pixels and then extracted a joint color/texture feature vector from each block. Here, the texture features are represented as the means and standard deviations of the coefficients of a bank of Gabor filters (with 3 scales and 4 orientations), and the color features are the mean values of HSV color components. Finally, for each image data set, we performed k -means clustering on the extracted feature vectors to form a vocabulary of 400 visual keywords. Since all of the constraint propagation algorithms finally apply spectral clustering to form the clusters, and the spectral clustering is a graph-based algorithm, based on this visual vocabulary, we then defined a spatial Markov kernel [33] as the similarity matrix W for graph construction. Moreover, to model local neighborhood relationships between the data points, we constructed a k -nearest neighbor graph, in which $w_{ij} = 0$ if x_j was not among the k -nearest neighbors (k -NN) of x_i . The k -NN graph construction was necessary for the spectral clustering [29]. In addition, we fixed $k = 20$ in all experiments.

In order to evaluate these algorithms, we compared the clustering results with the available ground-truth data labels, and employed normalized mutual information (NMI) as the performance measure [34]. For two random variables X and Y , NMI is defined as

$$\text{NMI}(X, Y) = \frac{I(X, Y)}{\sqrt{H(X)H(Y)}}$$

where $I(X, Y)$ is the mutual information between X and Y , and $H(X)$ and $H(Y)$ are the entropies of

Fig. 3 Sample images from 8 categories of the Scene data set and 15 categories of the Corel data set.

X and Y respectively. Note that $0 \leq \text{NMI} \leq 1$, and the larger NMI is, the better is the result. To evaluate the algorithms under different settings of pairwise constraints, we applied the ground-truth data labels to generate a varying number of pairwise constraints randomly for each data set. We randomly chose a pair of data points from each data set. If they had the same class labels, we generated a must-link constraint, otherwise a cannot-link constraint. In the experiments, we ran these algorithms 20 times with random initialization, and we report the average NMI (avg), the minimal NMI (min), and the standard variation (std).

4.2 Results on Scene data set

We first chose the Scene image data set to test the proposed algorithm. In the experiments, we compared the clustering performance of the five algorithms with a varied number of pairwise constraints, from 0.5% to 4%. The clustering results are shown in Table 1, which shows that the proposed LMRPCP algorithm performs better than the other four algorithms in most cases. Whether there are fewer or more pairwise constraints, our LMRPCP generally performs well. The effectiveness of our propagation method is verified by the fact that our LMRPCP consistently obtains better results in the constrained clustering task. In contrast, joint-MC performs unsatisfactorily, and in the case of few pairwise constraints, its performance is even worse than that of NCuts. This may be because when there are fewer pairwise constraints, the noise in the data matrix will heavily influence the matrix completion process and degrade its performance. On the other hand, as the number of constraints grows, the performance of our LMRPCP algorithm improves

more rapidly than that of AP and E²CP: our method can propagate the pairwise constraints more effectively than AP and E²CP.

Moreover, another important observation is that our LMRPCP performs very stably under different settings of pairwise constraints. However, the other three constraint propagation methods are less stable (measured by the standard variation). Note that as the pairwise constraints increase, the standard variation of E²CP grows, because E²CP only considers the local structure in the propagation process which therefore inevitably leads to the over-propagation phenomenon. With fewer pairwise constraints, joint-MC also has big variations due to data noise.

4.3 Results on the Corel data set

We also tested our proposed method on the Corel image data set. Clustering results are shown in Table 2, from which we may draw the same conclusions as for the Scene data set. Generally, LMRPCP performs better and more stably than the other three constraint propagation methods. This is because LMRPCP can not only consider the global low-rank structure but also remove the influence of the data noise.

To make it clearer how our LMRPCP exploits the pairwise constraints for spectral clustering, in Fig. 4, we show the distance matrices of the low-dimensional data representations obtained by NCuts, AP, E²CP, joint-MC, and LMRPCP when there are 1% pairwise constraints. We can see that the block structure of the distance matrices of the data representations obtained by LMRPCP on the two image data sets is significantly more obvious compared to those of the data representations obtained by NCuts,

Table 1 Clustering results (NMI) on Scene image data set under different settings of pairwise constraints, from 0.5% to 4%. The NCuts result is $\text{avg} = 0.54$, $\text{min} = 0.53$, and $\text{std} = 0.01$.

		0.5%	1.0%	1.5%	2.0%	2.5%	3.0%	3.5%	4.0%
AP	avg	0.57 ± 0.02	0.59 ± 0.02	0.62 ± 0.03	0.63 ± 0.02	0.63 ± 0.02	0.63 ± 0.02	0.64 ± 0.02	0.61 ± 0.06
	min	0.53	0.55	0.56	0.58	0.59	0.59	0.61	0.43
E ² CP	avg	0.70 ± 0.05	0.75 ± 0.05	0.81 ± 0.25	0.85 ± 0.20	0.83 ± 0.26	0.93 ± 0.16	0.91 ± 0.20	0.90 ± 0.14
	min	0.54	0.59	0.18	0.30	0.06	0.28	0.22	0.54
Joint-MC	avg	0.14 ± 0.15	0.82 ± 0.28	0.99 ± 0.02	0.97 ± 0.03	0.99 ± 0.02	0.99 ± 0.02	0.99 ± 0.02	0.98 ± 0.03
	min	0.01	0.13	0.94	0.94	0.93	0.94	0.94	0.93
LMRPCP	avg	0.78 ± 0.01	0.98 ± 0.00	1.00 ± 0.00	1.00 ± 0.01	0.99 ± 0.02	0.98 ± 0.02	0.96 ± 0.04	0.97 ± 0.03
	min	0.77	0.97	1.00	0.96	0.96	0.96	0.86	0.93

Table 2 Clustering results (NMI) on the Corel image data set under different settings of pairwise constraints, from 0.5% to 4%. The NCuts result is $\text{avg} = 0.51$, $\text{min} = 0.50$, and $\text{std} = 0.01$.

		0.5%	1.0%	1.5%	2.0%	2.5%	3.0%	3.5%	4.0%
AP	avg	0.50 ± 0.02	0.62 ± 0.02	0.61 ± 0.04	0.61 ± 0.04	0.59 ± 0.04	0.60 ± 0.03	0.59 ± 0.05	0.61 ± 0.03
	min	0.36	0.58	0.54	0.49	0.50	0.54	0.46	0.51
E ² CP	avg	0.63 ± 0.04	0.64 ± 0.02	0.76 ± 0.09	0.85 ± 0.06	0.82 ± 0.12	0.80 ± 0.13	0.83 ± 0.12	0.80 ± 0.11
	min	0.49	0.59	0.53	0.70	0.46	0.44	0.44	0.55
Joint-MC	avg	0.21 ± 0.12	0.61 ± 0.10	0.73 ± 0.22	0.86 ± 0.13	0.96 ± 0.01	0.96 ± 0.02	0.97 ± 0.01	0.96 ± 0.01
	min	0.03	0.32	0.08	0.36	0.94	0.92	0.94	0.95
LMRPCP	avg	0.57 ± 0.01	0.69 ± 0.01	0.81 ± 0.02	0.92 ± 0.02	0.97 ± 0.01	0.97 ± 0.01	0.98 ± 0.02	0.98 ± 0.01
	min	0.56	0.66	0.78	0.89	0.95	0.95	0.95	0.97

Fig. 4 Distance matrices of the low-dimensional representations for the two image data sets obtained by NCuts, AP, E²CP, joint-MC, and LMRPCP. The darker a pixel, the smaller the distance.

AP, E²CP, and joint-MC: after incorporating the pairwise constraints produced by our method, each cluster associated with the new data representation becomes more compact and different clusters become more separate. We can conclude that LMRPCP is effective as the pairwise constraints produced by LMRPCP can correctly reflect the category relationships between the data points.

5 Conclusions

In this paper, we have proposed a novel constraint

propagation approach, which is based on the low-rank nature of the propagation problem. We propose to make the clean data matrix and the constraint matrix jointly low-rank to remove data noise. Our approach is formulated as a low-rank matrix recovery (LMR) problem, which can be effectively solved with a modified augmented Lagrange multiplier (ALM) method. Experimental results on two real-life image data sets demonstrate the superiority of our proposed algorithm over the state-of-the-art.

Acknowledgements

The work described in this paper was supported by the National Natural Science Foundation of China (No. 61300164).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- [1] Basu, S.; Bilenko, M.; Mooney, R. J. A probabilistic framework for semi-supervised clustering. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 59–68, 2004.
- [2] Basu, S.; Davidson, I.; Wagstaff, K. *Constrained Clustering: Advances in Algorithms, Theory, and Applications*. Chapman & Hall/CRC, 2008.
- [3] Kamvar, S. D.; Klein, D.; Manning, C. D. Spectral learning. In: Proceedings of the 18th international joint conference on Artificial intelligence, 561–566, 2003.
- [4] Kulis, B.; Basu, S.; Dhillon, I.; Mooney, R. Semi-supervised graph clustering: A kernel approach. In: Proceedings of the 22nd international conference on Machine learning, 457–464, 2005.
- [5] Xing, E. P.; Jordan, M. I.; Russell, S. J.; Ng, A. Y. Distance metric learning with application to clustering with side-information. In: Advances in Neural Information Processing Systems 15, 2002. Available at <http://papers.nips.cc/paper/2164-distance-metric-learning-with-application-to-clustering-with-side-information.pdf>.
- [6] Li, Z.; Liu, J. Constrained clustering by spectral kernel learning. In: IEEE 12th International Conference on Computer Vision, 421–427, 2009.
- [7] Li, Z.; Liu, J.; Tang, X. Pairwise constraint propagation by semidefinite programming for semi-supervised classification. In: Proceedings of the 25th international conference on Machine learning, 576–583, 2008.
- [8] Lu, Z.; Ip, H. H. S. Constrained spectral clustering via exhaustive and efficient constraint propagation. *Lecture Notes in Computer Science* Vol. 6316, 1–14, 2010.
- [9] Lu, Z.; Carreira-Perpinán, M. A. Constrained spectral clustering through affinity propagation. In: IEEE Conference on Computer Vision and Pattern Recognition, 1–8, 2008.
- [10] Fu, Z.; Ip, H. H. S.; Lu, H.; Lu, Z. Multi-modal constraint propagation for heterogeneous image clustering. In: Proceedings of the 19th ACM international conference on Multimedia, 143–152, 2011.
- [11] Fu, Z.; Lu, H.; Ip, H. H. S.; Lu, Z. Modalities consensus for multi-modal constraint propagation. In: Proceedings of the 20th ACM international conference on Multimedia, 773–776, 2012.
- [12] Fu, Z.; Lu, H.; Li, W. Incremental visual objects clustering with the growing vocabulary tree. *Multimedia Tools and Applications* Vol. 56, No. 3, 535–552, 2012.
- [13] Fu, Z.; Lu, Z.; Ip, H. H. S.; Lu, H.; Wang, Y. Local similarity learning for pairwise constraint propagation. *Multimedia Tools and Applications* Vol. 74, No. 11, 3739–3758, 2015.
- [14] Fu, Z.; Lu, Z.; Ip, H. H. S.; Peng, Y.; Lu, H. Symmetric graph regularized constraint propagation. In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, 350–355, 2011.
- [15] Zhu, X.; Goldberg, A. B. Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* Vol. 3, No. 1, 1–130, 2009.
- [16] Cabral, R. S.; de la Torre, F.; Costeira, J. P.; Bernardino, A. Matrix completion for multi-label image classification. In: Advances in Neural Information Processing Systems 24, 2011. Available at <http://papers.nips.cc/paper/4419-matrix-completion-for-multi-label-image-classification.pdf>.
- [17] Goldberg, A.; Recht, B.; Xu, J.; Nowak, R.; Zhu, X. Transduction with matrix completion: Three birds with one stone. In: Advances in Neural Information Processing Systems 23, 2010. Available at <http://papers.nips.cc/paper/3932-transduction-with-matrix-completion-three-birds-with-one-stone.pdf>.
- [18] Candès, E. J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? *Journal of the ACM* Vol. 58, No. 3, Article No. 11, 2009.
- [19] Candès, E. J.; Recht, B. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* Vol. 9, No. 6, 717–772, 2009.
- [20] Fazel, M. Matrix rank minimization with applications. Ph.D. Thesis. Stanford University, 2002.
- [21] Candès, E. J.; Tao, T. The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory* Vol. 56, No. 5, 2053–2080, 2010.
- [22] Cai, J.-F.; Candès, E. J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* Vol. 20, No. 4, 1956–1982, 2010.
- [23] Ma, S.; Goldfarb, D.; Chen, L. Fixed point and Bregman iterative methods for matrix rank minimization. *Mathematical Programming* Vol. 128, Nos. 1–2, 321–353, 2011.
- [24] Lin, Z.; Chen, M.; Ma, Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv: 1009.5055, 2010.

- [25] Liu, G.; Lin, Z.; Yu, Y. Robust subspace segmentation by low-rank representation. In: Proceedings of the 27th International Conference on Machine Learning, 2010. Available at <http://www.icml2010.org/papers/521.pdf>.
- [26] Wright, J.; Ganesh, A.; Rao, S.; Peng, Y.; Ma, Y. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In: Advances in Neural Information Processing Systems 22, 2009. Available at <http://papers.nips.cc/paper/3704-robust-principal-component-analysis-exact-recovery-of-corrupted-low-rank-matrices-via-convex-optimization.pdf>.
- [27] Hale, E. T.; Yin, W.; Zhang, Y. Fixed-point continuation for l_1 -minimization: Methodology and convergence. *SIAM Journal on Optimization* Vol. 19, No. 3, 1107–1130, 2008.
- [28] Nocedal, J.; Wright, S. *Numerical Optimization*. New York, NY, USA: Springer, 1999.
- [29] Von Luxburg, U. A tutorial on spectral clustering. *Statistics and Computing* Vol. 17, No. 4, 395–416, 2007.
- [30] Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 22, No. 8, 888–905, 2000.
- [31] Oliva, A.; Torralba, A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* Vol. 42, No. 3, 145–175, 2001.
- [32] Bosch, A.; Zisserman, A.; Muñoz, X. Scene classification via pLSA. *Lecture Notes in Computer Science* Vol. 3954, 517–530, 2006.
- [33] Lu, Z.; Ip, H. H. S. Image categorization by learning with context and consistency. In: IEEE Conference on Computer Vision and Pattern Recognition, 2719–2726, 2009.
- [34] Strehl, A.; Ghosh, J. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research* Vol. 3, 583–617, 2003.

Zhenyong Fu received the Ph.D. degree in the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. He is currently an assistant professor in the College of Computer, Nanjing University of Posts and Telecommunications, China. His research interests include machine learning, computer vision, and multimedia processing.

Other papers from this open access journal are available at no cost from <http://www.springer.com/journal/41095>. To submit a manuscript, please go to <https://www.editorialmanager.com/cvmj>.